

(12) UK Patent Application (19) GB (11) 2 328 300 (13) A

(43) Date of A Publication 17.02.1999

(21) Application No 9814880.2

(22) Date of Filing 10.07.1998

(30) Priority Data

(31) 08893470 (32) 11.07.1997 (33) US

(71) Applicant(s)

International Business Machines Corporation
(Incorporated in USA - New York)
Armonk, New York 10504, United States of America

(72) Inventor(s)

Joe C Dubbels
Kevin P Gibson

(74) Agent and/or Address for Service

S R Davies
IBM United Kingdom Limited, Intellectual Property
Dept, Hursley Park, Winchester, Hampshire,
SO21 2JN, United Kingdom

(51) INT CL⁶

G06F 17/30

(52) UK CL (Edition Q)

G4A AADB

(56) Documents Cited

GB 2305525 A EP 0851367 A1
Seybold Report on Internet Publishing Vol. 2 No. 1,
September 1997, pages 5-20 & IAC Acc No 19756197

(58) Field of Search

UK CL (Edition P) G4A AADB
INT CL⁶ G06F 17/30
Online: WPI, EDOC, INSPEC, COMPUTER

(54) Abstract Title

Printing pages related to a viewed WWW page

(57) Enabling a user to select a web page, then print some or all of the linked pages which, based on predetermined criteria, are related to that page, without having first to invoke the linked pages. Each web page includes an applet (310, fig. 3) that runs on the web client system (200, fig. 3) when a print button on the page is pressed, 420. A print tool (330, fig. 3) running on the server (220, fig. 3) then parses the selected page and builds a list of related pages, 430, and allows the user to select which of the related pages will be printed, 440, 450. The print tool then constructs a temporary web page containing all the selected web pages, 460. This temporary web page is printed, 480, using the browser standard print function (320, fig. 3). Alternatively, a print utility in the web client performs the parsing and list building functions (figs. 5 and 6).

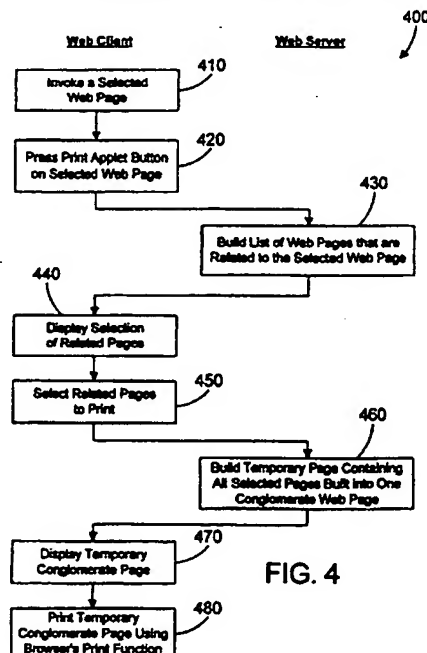


FIG. 4

This print takes account of replacement documents submitted after the date of filing to enable the application to comply with the formal requirements of the Patents Rules 1995

GB 2 328 300 A

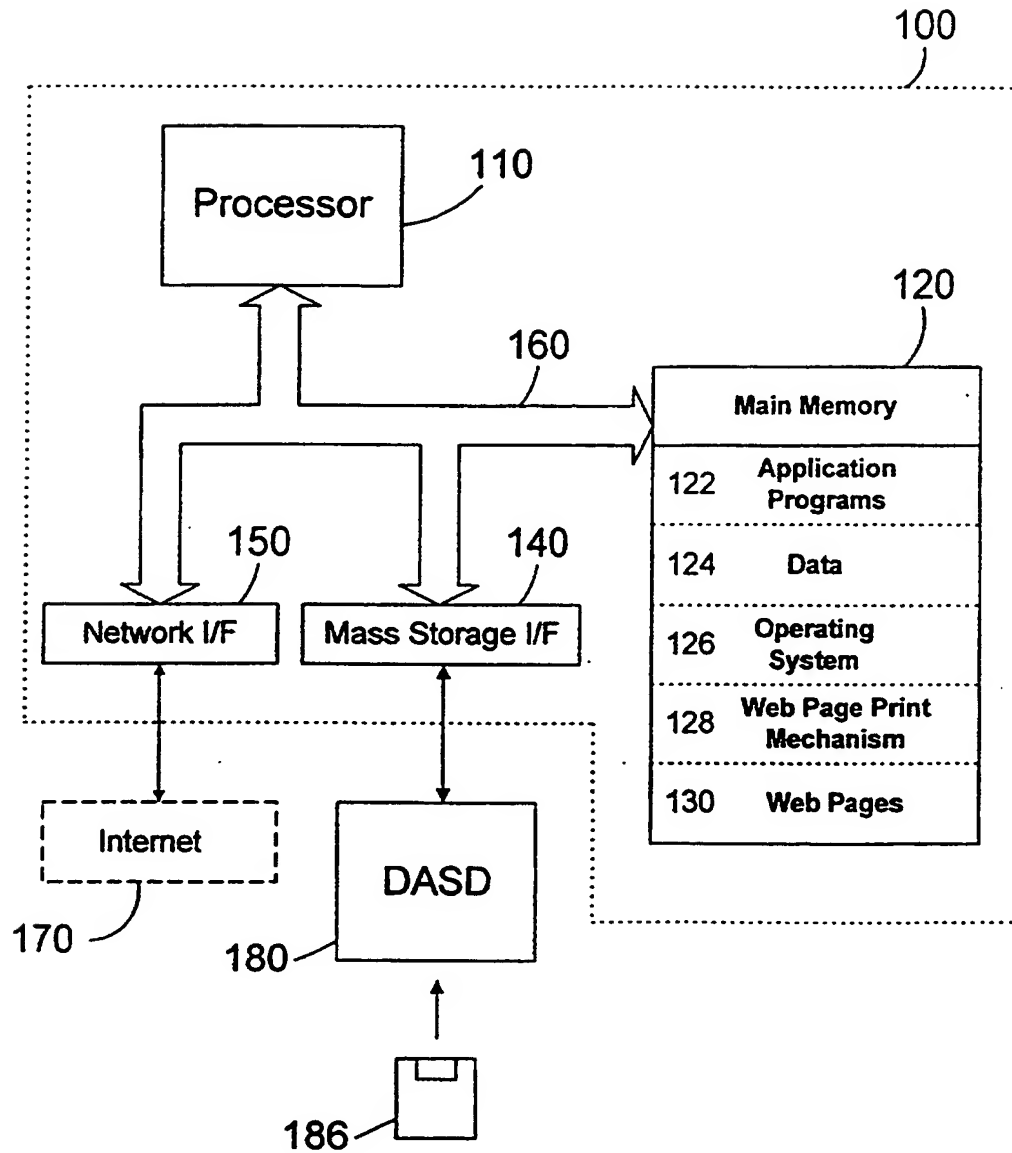


FIG. 1

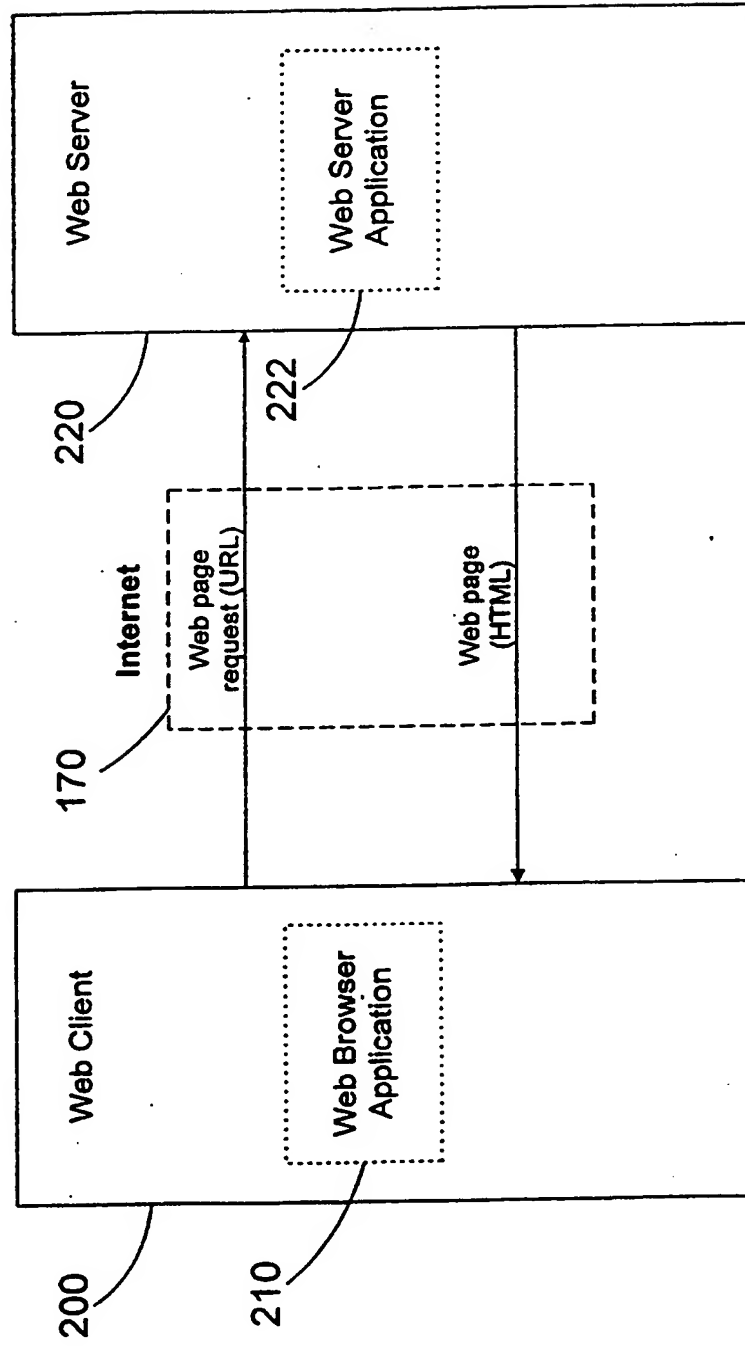


FIG. 2

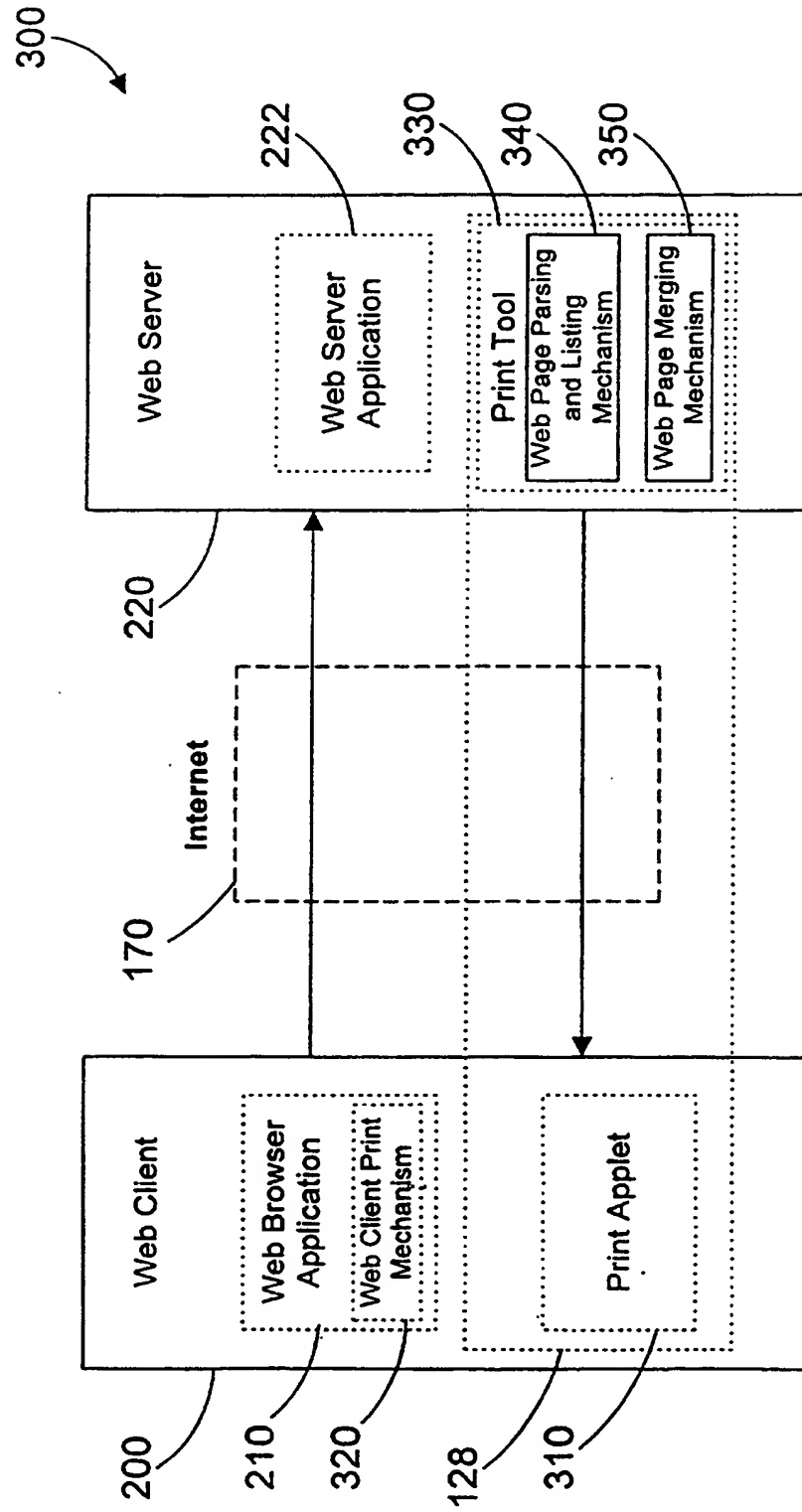


FIG. 3

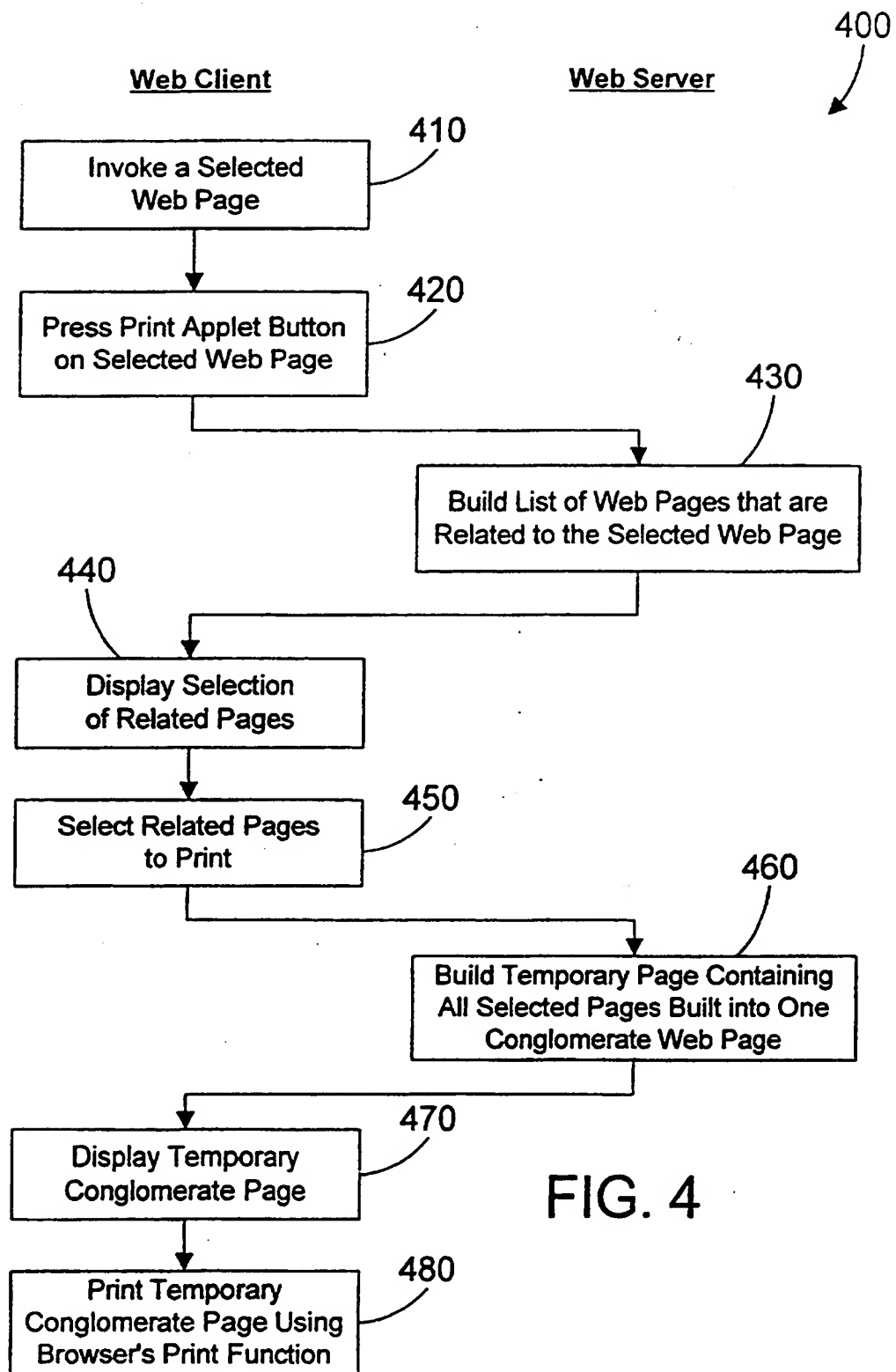


FIG. 4

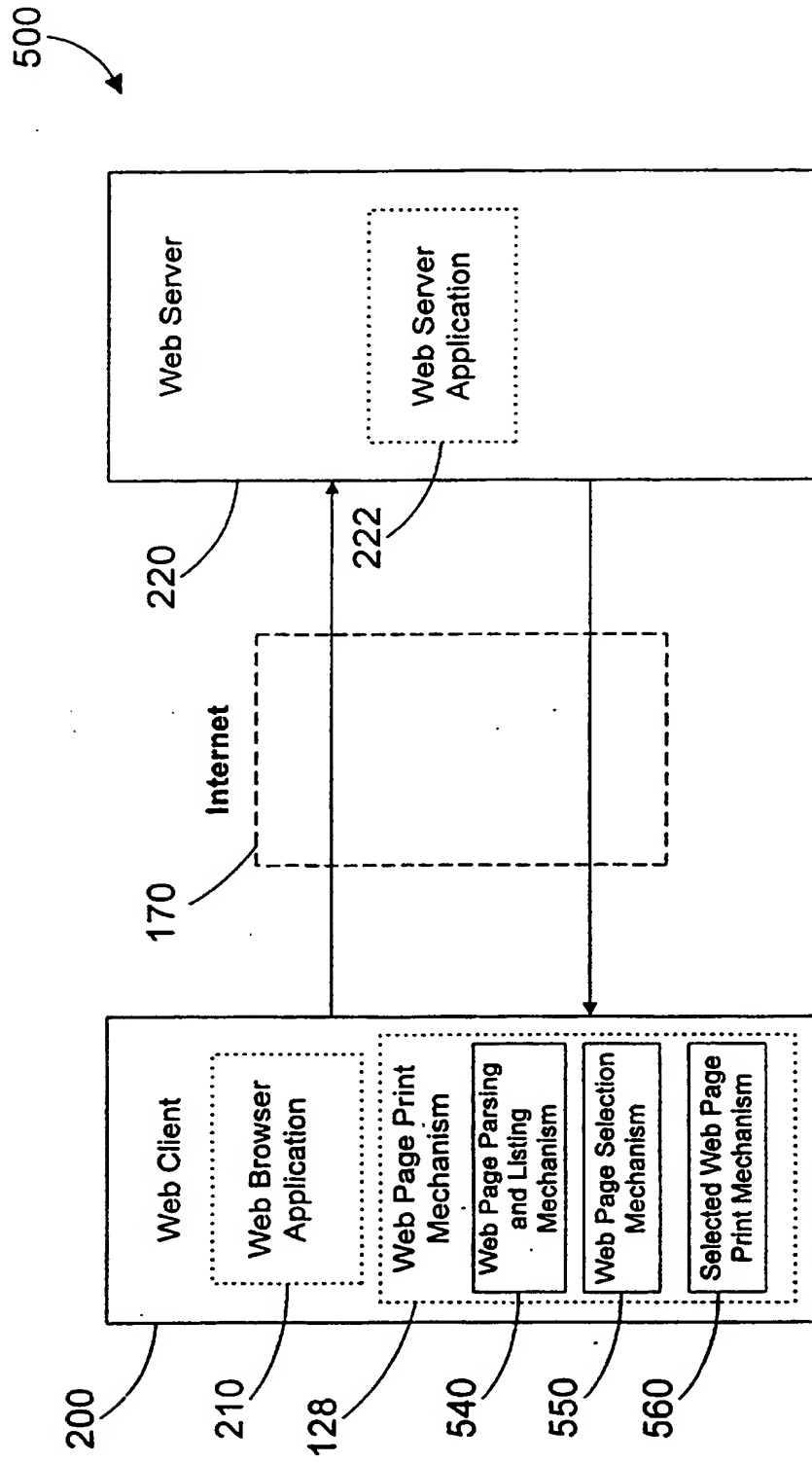


FIG. 5

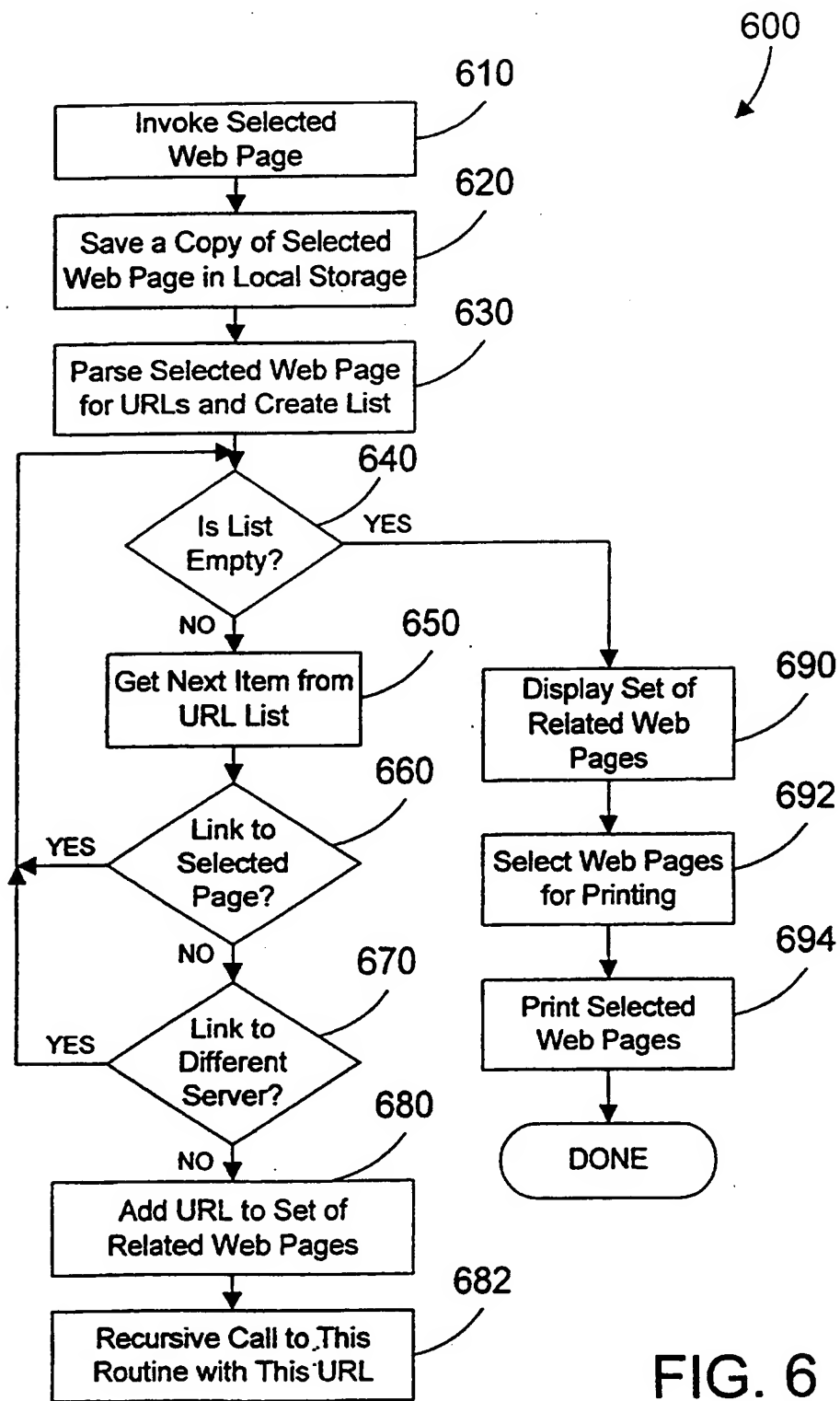


FIG. 6

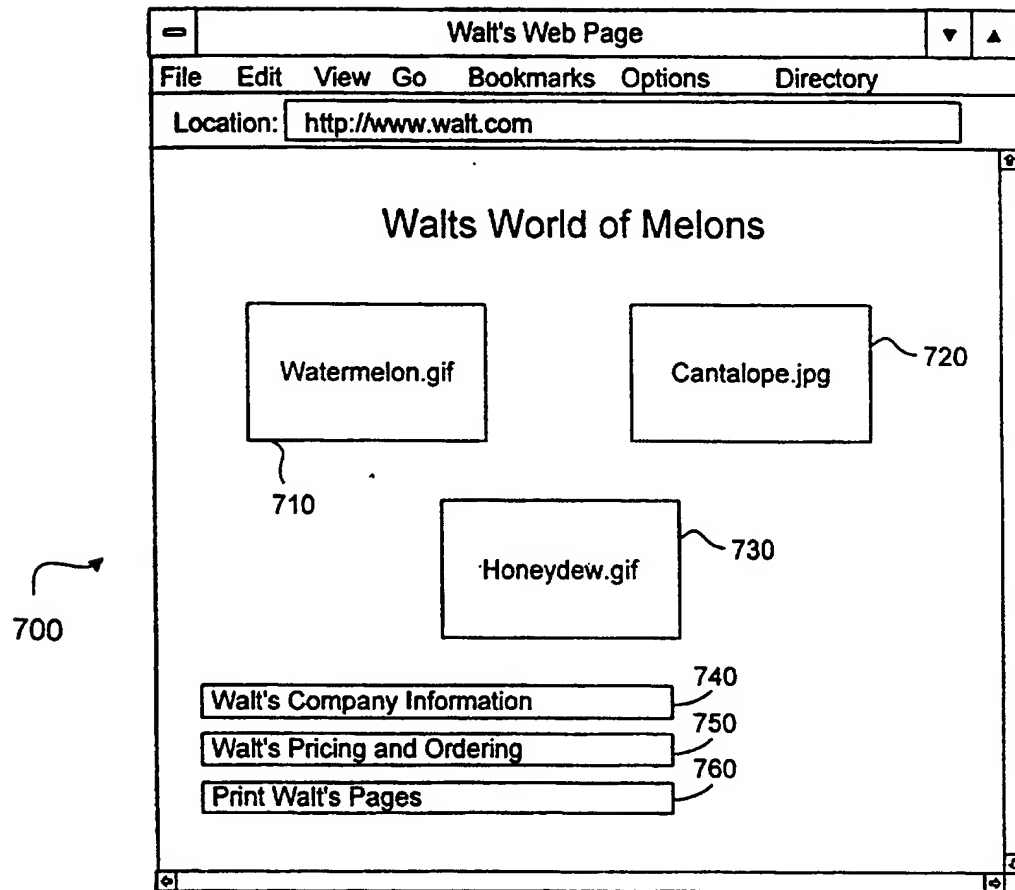


FIG. 7

800

810

	Web Page URL	Web Page Title
<input checked="" type="checkbox"/>	www.walt.com	Walt's World of Melons
<input type="checkbox"/>	www.walt.com\company	Walt's Company Information
<input checked="" type="checkbox"/>	www.walt.com\pricing	Walt's Pricing and Ordering
<input checked="" type="checkbox"/>	www.walt.com\delivery	Walt's Delivery Terms and Conditions

820

Print Checked Pages

FIG. 8

APPARATUS AND METHOD FOR PRINTING

This invention relates to an apparatus and method for printing material such as Web pages from computer networks such as the Internet.

The development of the EDVAC computer system of 1948 is often cited as the beginning of the computer era. Since that time, computer systems have evolved into extremely sophisticated devices, and computer systems may be found in many different settings. The widespread proliferation of computers prompted the development of computer networks that allow computers to communicate with each other. With the introduction of the personal computer (PC), computing became accessible to large numbers of people. Networks for personal computers were developed that allow individual users to communicate with each other.

One significant computer network that has recently become very popular is the Internet. The Internet grew out of this proliferation of computers and networks, and has evolved into a sophisticated worldwide network of computer systems. A user at an individual PC (*i.e.*, workstation) that wishes to access the Internet typically does so using a software application known as a web browser. A web browser makes a connection via the Internet to other computers known as web servers, and receives information from the web servers that is displayed on the user's workstation. Information displayed to the user is typically organized into pages (known as web pages) that are constructed using a specialized language called Hypertext Markup Language (HTML). Many web pages include one or more special reference locations known as "links" that invoke other web pages. Links allow a web user to easily navigate to other web sites of interest by clicking on the appropriate link with a mouse or other pointing device.

Often a web user will want to print a web page being currently viewed. Web browsers typically have a print function that allows a user to print the current page. However, as the complexity of web sites increases, it becomes increasingly difficult to locate needed information, and the process of printing several related web pages becomes a tedious exercise that involves: invoking the web page, printing the web page, invoking the next web page, printing, invoking, printing, etc. In other words, prior art browsers require a user to invoke a page before printing it. With these prior art browsers, if a user needs to print 40 related web pages, the user must manually invoke

and print each of the 40 web pages. Needless to say, this process becomes very time-consuming. As the number of Internet users, providers, and web servers continues to rapidly expand, such problems in the printing of web pages will continue to be an impediment to the effective usage of resources available on the Internet.

Accordingly, the invention provides apparatus for printing multiple pages comprising:

means for selecting a page;

means for automatically determining one or more pages related to said selected page;

and means for printing said selected page and said one or more related pages.

Thus the convenience of printing related web pages or other suitable material is improved by providing ways that a user may print related pages without the customary user interaction required to invoke and print each page.

In a preferred embodiment, the apparatus further comprises a page parsing and listing mechanism, and the selected page is in hypertext markup language (HTML) and is selected using a Uniform Resource Locator (URL). A page can be considered as related to said selected page if they both reside on the same server, and optionally if they also both have the same base address, or using any other suitable criteria.

In one embodiment, said apparatus comprises a computer program product comprising computer program instructions recorded on a storage medium. In an alternative embodiment, said apparatus comprises a computer system including at least one processor, a memory coupled to the at least one processor; and computer program instructions in said memory for execution by the program to perform said printing operation. Moreover, it will be appreciated that the apparatus may comprise any suitable configuration of hardware and software, whether the apparatus is a single computer system or is comprised of multiple computer systems operating in concert.

The invention further provides apparatus including a web page print mechanism comprising:

a web page selection mechanism that allows a user to select at least one web page from the list of web pages;

a web page parsing and listing mechanism that generates a list of web pages related to the selected web page; and
and a mechanism for printing the at least one web page selected by the user using the web page selection mechanism.

5

The invention further provides a method for printing a plurality of pages, the method including the steps of:

selecting at least one page containing at least one reference to at least one other page;

10

parsing the at least one page to locate the at least one reference;

determining whether the at least one other page corresponding to the at least one reference is related to the selected page;

generating at least one list of pages related to the selected page;

and

15

printing a plurality of the pages in the at least one list.

Preferably the method further comprises the steps of:

displaying to a user the at least one list;

the user selecting from the at least one list which pages to print;

20

wherein the step of printing the plurality of web pages comprises the step of printing the user-selected pages.

Viewed from another aspect the invention provides apparatus comprising:

25

at least one processor;

a memory coupled to the at least one processor; and

a print mechanism residing in the memory and executed by the at least one processor, the print mechanism printing a plurality of pages that are related to a selected page.

30

In a preferred embodiment, the print mechanism further comprises a page parsing and listing mechanism, a page selection mechanism, and a mechanism for printing at least one page selected using the page selection mechanism.

35

Viewed from another aspect the invention further provides apparatus comprising:

at least one processor;

a memory coupled to the at least one processor;

40

a selected web page residing in the memory; and

a web page print mechanism residing in the memory and executed by the at least one processor, the web page print mechanism comprising:

a web page parsing and listing mechanism that generates a list of web pages related to the selected web page;

5 a web page selection mechanism that allows a user to select at least one web page from the list of web pages; and

a mechanism for printing the at least one web page selected by the user using the web page selection mechanism.

10 Viewed from another aspect the invention further provides a program product comprising:

(A) a print mechanism, the print mechanism printing a plurality of pages that are related to a selected page;

(B) signal bearing media bearing the print mechanism.

15

Such a program product may be distributed in a variety of forms, without detriment to operation in accordance with the present invention, regardless of the particular type of signal bearing media used to actually carry out the distribution. Examples of suitable signal bearing
20 media include: recordable type media such as floppy disks and CD ROM, and transmission type media such as digital and analog communications links.

In a preferred embodiment, the print mechanism comprises a web page print mechanism including:

25 a web page parsing and listing mechanism that generates a list of web pages related to a selected web page; and

a web page selection mechanism that allows a user to select from the list of web pages at least one web page to be printed.

30 Thus the apparatus and method described herein for printing related web pages allow a web user to select a web page, then print all of the related web pages based on one or more predetermined criteria. A web user is therefore able to print related web pages without manually invoking and printing each page. In a first embodiment, each web page
35 includes an applet that is run on the web client system when a print button on the page is pressed. The client applet communicates with a print tool running on the server that parses the selected page and builds a list of related pages and allows the user to select which of the related pages will be printed. Once the user selects the pages to be
40 printed, the print tool constructs a temporary web page that contains all the web pages the user selected. This temporary web page may then be

printed using the standard print function supplied with the browser. In a second embodiment, a print utility in the web client allows a user to print related web pages by parsing a selected web page and building a list of related pages. The user may then select from the list of related pages which pages to print. The selected pages are then printed.

Preferred embodiments of the invention will now be described in detail by way of example only with reference to the following drawings:

FIG. 1 is a block diagram of a computer system;

FIG. 2 is a block diagram of a typical Internet connection;

FIG. 3 is a block diagram of a computer system that allows printing of related web pages in accordance with a first embodiment of the invention;

FIG. 4 is a flow diagram of the method steps for printing related web pages in accordance with the first embodiment;

FIG. 5 is a block diagram of a computer system that allows printing of related web pages in accordance with a second embodiment;

FIG. 6 is a flow diagram of the method steps for printing related web pages following the second embodiment;

FIG. 7 is a sample web page; and

FIG. 8 is a sample display used to select web pages to be printed that relate to the web page of FIG. 7.

For those individuals who are not familiar with the Internet, a brief overview of relevant Internet concepts is first presented here. Thus an example of a typical Internet connection is shown in FIG. 2. A user that wishes to access information on the Internet typically has a computer workstation 200 that executes an application program known as a web browser 210. Under the control of web browser 210, workstation 200 sends a request for a web page over the Internet. Web page data can be in the form of text, graphics and other forms of information. Each web server on the Internet has a known address which the user must supply to the web browser in order to connect to the appropriate web server. Because web server 220 can contain more than one web page, the user will also specify in the address which particular web page he or she wants to view on web server 220. A web server computer system 220 executes a web server application 222, monitors requests, and services requests for which it has responsibility. When a request specifies web server 220, web server application 222 generally accesses a web page corresponding to the specific request, and transmits the page to the user's workstation 200.

A web page is primarily visual data that is intended to be displayed on the monitor of user workstation 200. Web pages are generally written in Hypertext Markup Language (HTML). When web server 220 receives a web page request, it will build a web page in HTML and send it off across the Internet 170 to the requesting web browser 210. Web browser 210 understands HTML and interprets it and outputs the web page to the monitor of user workstation 200. This web page displayed on the user's screen may contain text, graphics, and links (which reference addresses of other web pages.) These other web pages (i.e., those represented by links) may be on the same or on different web servers. The user can go to these other web pages by clicking on these links using a mouse or other pointing device. This entire system of web pages with links to other web pages on other servers across the world is known as the "World Wide Web".

Referring now to FIG. 1, a computer system 100 includes a processor 110, a main memory 120, a mass storage interface 140, and a network interface 150, all connected by a system bus 160. Those skilled in the art will appreciate that this system encompasses all types of computer systems: personal computers, midrange computers, mainframes, etc. Note that many additions, modifications, and deletions can be made to this computer system 100; examples of possible additions include: a computer monitor, a keyboard, a cache memory, and peripheral devices such as printers.

Processor 110 can be constructed from one or more microprocessors and/or integrated circuits. Processor 110 executes program instructions stored in main memory 120. Main memory 120 stores programs and data that the computer may access. When computer system 100 starts up, processor 110 initially executes the program instructions that make up operating system 126. Operating system 126 is a sophisticated program that manages the resources of the computer system 100. Some of these resources are the processor 110, main memory 120, mass storage interface 140, network interface 150, and system bus 160.

Main memory 120 includes one or more application programs 122, data 124, operating system 126, a web page print mechanism 128, and one or more web pages 130. Application programs 122 are executed by processor 110 under the control of operating system 126. Application programs 122 can be run with program data 124 as input. Application programs 122 can also output their results as program data 124 in main memory. As

described herein, the computer system 100 includes a web page print mechanism 128 that allows multiple related web pages to be printed without manually printing each web page. The web page print mechanism 128 of FIG. 1 may exist on a single computer system, as shown in FIG. 5, or may be distributed among multiple computer systems, as shown in FIG. 3.

Mass storage interface 140 allows computer system 100 to retrieve and store data from auxiliary storage devices such as magnetic disks (hard disks, diskettes) and optical disks (CD-ROM). These mass storage devices are commonly known as Direct Access Storage Devices (DASD), and act as a permanent store of information. One suitable type of DASD is a floppy disk drive 180 that reads data from and writes data to a floppy diskette 186. The information from the DASD can be in many forms. Common forms are application programs and program data. Data retrieved through mass storage interface 140 is usually placed in main memory 120 where processor 110 can process it.

While main memory 120 and DASD device 180 are typically separate storage devices, computer system 100 uses well known virtual addressing mechanisms that allow the programs of computer system 100 to behave as if they only have access to a large, single storage entity, instead of access to multiple, smaller storage entities (e.g., main memory 120 and DASD device 185). Therefore, while certain elements are shown to reside in main memory 120, those skilled in the art will recognize that these are not necessarily all completely contained in main memory 120 at the same time. It should be noted that the term "memory" is used herein to generically refer to the entire virtual memory of computer system 100.

Network interface 150 allows computer system 100 to send and receive data to and from any suitable network to which the computer system may be connected. This network may be a local area network (LAN), a wide area network (WAN), or more specifically the Internet 170. Suitable methods of connecting to the Internet include known analog and/or digital techniques. Many different network protocols can be used to implement a network. These protocols are specialized computer programs that allow computers to communicate across a network. TCP/IP (Transmission Control Protocol/Internet Protocol), used to communicate across the Internet, is an example of a suitable network protocol.

System bus 160 allows data to be transferred among the various components of computer system 100. Although computer system 100 is shown to contain only a single main processor and a single system bus, those skilled in the art will appreciate that embodiments may be formed using a computer system that has multiple processors and/or multiple buses. In addition, the interfaces that are used in the preferred embodiment may include separate, fully programmed microprocessors that are used to off-load compute-intensive processing from processor 110, or may include I/O adapters to perform similar functions.

Two different embodiments for printing related web pages will now be described. The first embodiment uses an applet on the web client in conjunction with a print tool that resides on the web server. An example of a suitable system and method in accordance with the first embodiment is shown in FIGS. 3 and 4. Any of the programs executing on a web server are referred to generically herein as web server programs, and any of the programs executing on the web client are referred to generically herein as web client programs. The second embodiment of the present invention does not require any software to be installed on the web server. An example of a suitable system in accordance with the second embodiment is shown in FIGS. 5 and 6.

An apparatus 300 in accordance with the first embodiment is illustrated in FIG. 3, and includes a web client 200 coupled to a web server 220 via the Internet 170. Web client 200 includes a web browser application 210 and a print applet 310. The function of web browser application 210 is described above and is well-known in the art, and includes a web client print mechanism 320 that is used to print individual web pages. Print applet 310 is a small application such as a Java applet that is invoked when a user takes a particular action with respect to a selected web page. In the preferred embodiment, print applet 310 is executed when a user selects a particular "print button" on a web page that includes print applet 310. Print applet 310 is shown to reside on web client 200, but those skilled in the art will recognize that applets such as print applet 310 are typically dynamically loaded from web server 220 to web client 200 with a web page. A user presses the "print button" on a web page that corresponds to print applet 310 to indicate that printing of the current page and its related pages is desired.

Web server 220 includes a web server application 222 and a print tool program 330. Print tool 330 is a web server program that is used in conjunction with print applet 310 to print multiple related web pages. Print tool 330 includes a web page parsing and linking mechanism 340 and a web page merging mechanism 350. The function of mechanisms 340 and 350 may best be understood with reference to a method 400 in accordance with the first embodiment, which is illustrated in FIG. 4.

Method 400 starts by the web client 200 invoking a selected web page (step 410). A web page is typically invoked by the web client sending a Uniform Resource Locator (URL) to web server 220. The user can send a URL by "clicking" with a mouse on a web page link, or the user can enter the entire URL address manually in the web browser. The URL is sent and travels across the Internet 170, contacting the web server 220 that is specified in the URL. The web server then delivers the requested page specified by the URL to the web client. Note that the process of invoking a selected web page in step 410 involves web client/web server interaction, which is omitted from FIG. 4 for the sake of clarity. The mechanisms for invoking a web page using a URL are well-known and understood in the art.

Once the selected page is displayed on the web client, the user may press the print applet button on the web page (step 420) to cause the web page with its related web pages to be printed. Print applet 310 communicates with print tool 330 in web server 220, and in response, the web page parsing and listing mechanism 340 in print tool 330 builds a list of web pages that are related to the selected web page (step 430). Web page parsing and listing mechanism 340 builds the list by first parsing the selected web page and examining all links in the selected web page. Each of these links are analyzed to determine whether or not the link points to a web page that is related to the selected web page. Any suitable criteria can be used to determine whether or not two web pages are "related". For example, one suitable criteria for relating web pages determines that web pages on the same web server are related, while pages on different web servers are not related. A preferred criteria for relating web pages determines that web pages that have the same base address as part of their URL are related, while pages with different base addresses are not. For example, a home page may have the address `www.companyX.com/home.html`, and any pages that have the base address `www.companyX.com` are related to the home page. In another example, a page at address `www.companyX.com/support/index.html` is selected, and any

pages that share the base address `www.companyX.com/support` are related to the selected page, while other pages at this site are not related. Regardless of the specific criteria used, pages that are related to the selected web page are included in the list, and pages that are not related are not included in the list.

5

The list of related web pages is then passed to the web client, which displays the list to the user (step 440). The user then selects the pages on the list to print (step 450). The list of selected pages is then passed to the web server, which uses this information to build a temporary web page that is a conglomerate of all the pages that were selected for printing (step 460). The temporary conglomerate web page is built by the web page merging mechanism 350, which performs the necessary functions to convert several individual web pages into a single web page. For example, a tag `<body>` generally defines the beginning of an HTML page, and the tag `</body>` defines the end of an HTML page. For the case of printing HTML pages, web page merging mechanism 350 builds the conglomerate web page by removing the `</body>` tag in the first page to be printed, by removing the `<body>` tag in the last page to be printed, and by removing all `<body>` and `</body>` tags for all pages in between. In addition, other tags such as header and end tags may be moved to the beginning or end of the conglomerate web page, or may be deleted, if appropriate. This results in a single conglomerate web page that contains all the pages to be printed. This conglomerate web page is then passed to the web client and displayed to the user (step 470). The conglomerate web page may then be printed using the conventional print function that is supplied with the web browser application (step 480).

10

15

20

25

By providing a print applet that is downloaded to a web browser with a web page along with a print tool program running on the web server, a user may print multiple related pages with a standard web browser. This approach requires new software to be added to the web server. In the second embodiment, discussed in more detail below, no additional software is added to the web server. Instead, software is added to the web client to provide the capability of printing multiple related web pages.

30

35

Referring now to FIGS. 5 and 6, an apparatus 500 in accordance with the second embodiment includes a web client 200 and a web server 220 connected via the Internet 170. Web client 200 includes a web browser application 210 and a web page print mechanism 128. The web browser

40

application 210 is a standard web browser known in the art. Web page print mechanism 128 includes a web page parsing and listing mechanism 540, a web page selection mechanism 550, and a selected web page print mechanism 560. While web page print mechanism 128 is shown in FIG. 5 as being separate from web browser 210, it is contemplated that web page print mechanism 128 will be integrated into a web browser application, thereby providing a browser with advanced web page printing capability. In the alternative, web page print mechanism 128 may be a separate application running on web client 200, or may be a plug-in or Java applet/application for web browser application 210. The functions of web page print mechanism 128 are described herein without regard to whether mechanism 128 resides within web browser application 210 or outside of web browser application 210.

Web page parsing and listing mechanism 540 is used to create a list of related web pages. Web page selection mechanism 550 interfaces with the web user to allow the user to select which of the related web pages in the list to print. Selected web page print mechanism 560 takes the web pages selected by the user in the list of related web pages and prints them. The function of these mechanisms may best be understood with relation to the flow diagram of FIG. 6.

A method 600 for printing multiple related web pages begins by invoking a selected web page (step 610). As discussed above with reference to the first embodiment, the mechanisms and interplay between web client and web server to invoke a web page are well-known in the art, and are not discussed here. Once the selected web page has been invoked, a copy of the selected web page is saved in local storage (step 620). Local storage may include any portion of memory within apparatus 500, including main memory, DASD, or other storage devices. The selected web page is then parsed by web page parsing and listing mechanism 540 for links to other web pages and a list of these links is created (step 630). Note that web page parsing and listing mechanism 540 performs steps 640-682 described below. The list is processed beginning with step 640. If the list is not empty (step 640=NO), the next URL in the list is selected (step 650). If the URL is a link to the selected page (step 660=YES) or is a link to a different server (step 670=YES), the URL is ignored and not added to the set of related web pages. If the URL is to a related page (e.g., on the same server for this example) (steps 660 and 670=NO), the URL is added to the set of related web pages (step 680), and method 600 is invoked recursively for that URL. In this manner method 600